**Large-scale discovery of visual features for object recognition**
D. Linsley, S. Eberhardt, & T. Serre

A central goal in vision science is to identify features that the visual system uses to recognize objects and scenes. Reverse correlation methods have been used to uncover features important for recognizing faces and other stimuli with low intra-class variability. However, these methods are less successful when applied to natural scenes that exhibit variability in their appearance.

To rectify this, we developed Clicktionary, a web-based game used for measuring visual feature importance for recognizing real-world objects. Pairs of participants play together in different roles to identify objects: A "teacher" reveals image regions diagnostic of the object's category while a "student" tries to recognize the object as quickly as possible. Aggregating game data across players yields importance maps for individual object images, in which each pixel is scored by its contribution to object recognition. We found that these importance maps are consistent across participants and identify object features that are distinct from those used by state-of-the-art deep convolutional networks (DCNs) for object recognition or those predicted by salience maps derived from both human participants and models.

We also extended Clicktionary to support large-scale feature map discovery (http://clickme.ai), whereby human teachers play with DCN students. To date we have gathered a dataset of over several tens of thousands of unique images, which is large enough to incorporate into DCN training routines and begin teaching them to recognize objects using similar visual features as humans. Indeed, by "cuing" DCNs during object recognition training to base their decisions on features emphasized in Clicktionary maps, we have significantly altered their object representations, reducing the reliance on background information and imbuing them with feature importance maps that are more similar to humans.

In summary, we present a new procedure for measuring feature importance for object recognition in humans, and a novel method for incorporating those features into vision models. Human feature importance maps identified by Clicktionary and our DCN models trained with this information will enable a richer understanding of the foundations of object recognition.

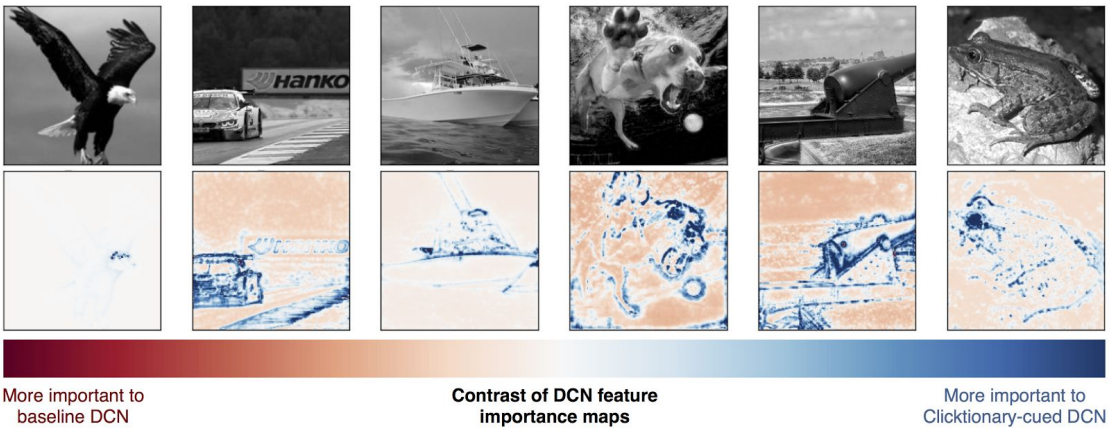


**Figure.** DCNs cued to emphasize Clicktionary feature importance maps become less reliant on background features. We developed a procedure for DCN object classification training whereby they are cued to emphasize feature importance maps derived from Clicktionary. This significantly alters learned representations in the "Clicktionary-cued" DCNs, de-emphasizing background information in exchange for object parts that are also diagnostic for human participants. This difference is plotted for selected images held-out of DCN training as the contrast between standardized feature importance maps derived from a cued DCN versus an uncued one ([Clicktionary-cued feature importance] - [baseline DCN]).

**Funding**: This work is supported by NSF early career award (IIS-1252951) and DARPA young faculty award (N66001-14-1-4037) to TS. DL is also partly supported by Brown Institute for Brain Sciences (BIBS).